Personality Homophily and Geographic Distance in Facebook

Nyala Noë[1,2], Roger M. Whitaker[1], and Stuart M. Allen[1]

[1] *School of Computer Science and Informatics*

Cardiff University

Queen's Building, 5 The Parade, Roath

Cardiff CF24 3AA, United Kingdom

[2] *Corresponding author: noen@cardiff.ac.uk*

Abstract

Personality homophily remains an understudied aspect of social networks, with the traditional focus concerning socio-demographic variables as the basis for assortativity, rather than psychological dispositions. We consider the effect of personality homophily on one of the biggest constraints to human social networks: geographic distance. We use the Big five model of personality to make predictions for each of the five facets: Openness to experience, Conscientiousness, Agreeableness, Extraversion, and Neuroticism. Using a network of 313,669 Facebook users, we investigate the difference in geographic distance between homophilous pairs, in which both users scored similarly on a particular facet, and mixed pairs. In accordance with our hypotheses, we find that pairs of open and conscientious users are geographically further apart than mixed pairs. Pairs of extraverts, on the other hand, tend to be geographically closer together. We find mixed results for the Neuroticism facet, and no significant effects for the Agreeableness facet. The results are discussed in the context of personality homophily and the impact of geographic distance on social connections.

***Keywords*:** personality homophily, geographic distance, Big Five, Facebook.

Introduction

People have traditionally been socially and geographically constrained in their choice of friendships, frequently establishing new ties through mutual acquaintances or shared activities.[1] Furthermore, people have a tendency to form connections to others with whom they are similar [2,3], making such relationships easier to support. This tendency to assort based on similarity is known as *homophily*. The similarity that drives homophily may occur from a range of socio-demographic variables, such as religion, age, race, gender, and level of education. [3] However, studies of homophily based on people's psychological dispositions, such as personality, remain sparse in the literature. Personality is a crucial precursor of human behavior and can explain why people behave differently in the same situation.[4] The Five Factor model of personality[5], or the Big 5, has been widely used and validated.[6] This model of personality consists of five facets: Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Openness to experience characterizes people who are curious, creative, and willing to try new experiences. Conscientious people are organized, timely, and thoughtful towards others. Extraversion describes our sociability and willingness to meet new people, while Agreeableness reflects how friendly we are to other people. Finally, Neuroticism describes how prone we are to worry and anxiety. The relationship between personality and social media use has been extensively discussed in a cross-cultural review paper by Zuniga and colleagues[7], but this topic is beyond the scope of this study.

Personality is an influential factor in offline social network structure.[11] Extraverts are known for their large social networks and their proneness to initiate new social contacts, while people who score high on the Agreeableness facet tend to be at the receiving end of social connections.[8,9,11] Open individuals tend to have diverse social networks, which might lead to

more sporadic connections and a network low in transitivity.[11,12] Conscientiousness is associated with the maintenance of personal ties, but appears to affect position in organizational networks more than personal ones: conscientious people are often at the center of their work networks.[11] Finally, Neuroticism does not seem to have a major impact on network structure, although it has been associated with more weakly connected ties, and tends to be negatively associated with centrality.[13]

Personality has been found to be homophilous in both offline[8] and online contexts.[9] In particular, three facets of the Five Factor Model stood out as homophilous in both contexts: Openness to experience, Extraversion, and Agreeableness.[8,9] Similarly, people connected in a Facebook friendship network were more similar in terms of Conscientiousness, Agreeableness, and Openness to experience, compared to users who were not connected.[10]

Given the Internet revolution and dominance of "virtual" online interactions, the extent of physical co-location has become an important issue to consider. The Internet and online social media provide an alternative platform for individuals to find common ground through shared interests, or similar attitudes and beliefs. Despite the opportunity that the Internet provides to create relationships without meeting, the importance of physical proximity in tie formation is just as strong in online social networks, as in offline ones.[14]

The probability of two people becoming friends is well-known to decrease with geographic distance [15–19]. Distance appears to matter less when people are close enough to easily be able to travel to see each other face-to-face, or engage in activities together. [20] Communication networks appear to have two main levels: short-distance communication has high clustering, but is of short duration, while long distance communication has smaller clustering, but tend to last longer.[16]

Personality has been studied in spatial contexts at different levels of granularity.[21–23] Regions in the UK can be distinguished based on their differences in personality. For example, Scotland is high in Agreeableness, while Wales scores high on Neuroticism.[23] At the city level, differences in neighborhoods could be observed, with some scoring on average higher on Openness to experience and others higher in Agreeableness.[22] At the finest level of granularity lie the venues that individuals visit, where the perceived personality of patrons could be used to infer the ambiance of a venue.[21] Previous work has also shown that personality facets, such as Conscientiousness, Openness, and Neuroticism, are related to Foursquare usage.[24,25] Furthermore, Conscientiousness, Openness to experience and Agreeableness were found to be homophilous, using Foursquare users' check-in history. People who scored highly on each of these facets were more likely to go to the same locations as similar others.[26] It is therefore important to consider the influence these spatial influences can have on the formation, shape, and evolution of social networks.

In this paper, we focus on the interaction of geographic distance and the different facets of the Five Factor model of personality on the connectedness of people in an online social network. Studies on personality homophily rarely consider one of the constraints to the formation and maintenance of ties: geographic distance.

Motivation for Hypotheses

We hypothesize that people with certain personality traits are more affected by distance than others. Open people might be more likely to be connected to people further away, as they are more likely to initiate contacts with a range of different people. In contrast, we would expect more conservative people to be located geographically closer to each other (Hypothesis 1).

Conscientious people are organized and mindful of others[27], which might make it more likely for them to stay in contact, even if they are further away and face-to-face contact is not possible. We therefore expect the connection between fellow conscientious people to prevail even at longer distances, compared to pairs of friends who are low in Conscientiousness (Hypothesis 2).

We hypothesize that distance matters for Extraversion: people who are extraverts rely on physical, rather than online, activities to form friendships and maintain their social bonds.[28] Such activities require relative geographic proximity and therefore there is a basis for distance among extraverts to be lower compared to introverts (Hypothesis 3).

Agreeable people are popular friendship and communication partners, whether offline or online.[8,9] People might be motivated to stay in contact with others who are friendly. People who are high in Agreeableness might also be more likely to maintain contact with others. We therefore expect agreeable people to be connected, despite the distance (Hypothesis 4).

It is harder to make predictions for Neuroticism. Previous research suggests that neurotic people tend to have smaller groups of friends, and might use online interactions to substitute offline ones.[29] Neurotic people might also need emotional support that is readily available, making functional long distance relationships less likely (Hypothesis 5).

Summary of Hypotheses

1. Mutually open pairs tend to be further apart, while mutually not-open pairs tend to live in close proximity to one another, compared to mixed pairs.
2. Mutually conscientious pairs maintain friendships at greater distances, compared to mixed and mutually unconscientious pairs.

3. Mutually extraverted pairs tend to be in closer proximity to one another, compared to mutually introverted pairs and mixed pairs.
4. Mutually agreeable pairs are connected at greater distances, compared to disagreeable or mixed pairs.
5. Mutually neurotic pairs tend to maintain friendships at shorter distances, compared to emotionally stable or mixed pairs.

Methodology

We use a subset of variables from the MyPersonality triads dataset[30], which contains personality, geographic, and demographic information of 300,669 Facebook users (Table 1). From this triad dataset, we derive pairs of connected users, their personality, and geographic distance. The MyPersonality data is derived from Facebook users who answered a Personality questionnaire and provided access to their Facebook information and location. Data collection was carried out between June 2007 and 2012.[30]

Table 1. Description of variables used in this paper

variable	refers to	type	unit
personality tercile score	user	categorical	low middle high
personality similarity score	pair	categorical	same low same high mixed
distance	pair	continuous	kilometres

Geographic distance refers to the distance (in km) between two users in a pair, inferred from their latitude and longitude at the moment of completing the MyPersonality questionnaire. Distance and personality data was available for 289,557 pairs of users from the MyPersonality

dataset. 71,953 pairs of users had missing data for the distance measure, bringing the total number of valid pairs to 217,604.

The personality scores for all users were divided into terciles to obtain a categorical personality variable. For each facet, we consider low and high scorers, allowing a focus on the facet's extremes and clear-cut comparisons, such as between extraverts and introverts, for example (see Table 2). Middle scorers were taken out of the analyses; this approach has been successfully applied in previous work.[26,28,31]. This also allows for comparisons with previous studies with the same approach.[9] Finally, this also increases statistical sensitivity as the effect sizes in these type of studies tend to be small.[10] For each facet, this further brings down the sample size, as pairs with at least one middle scorer are taken out. For Openness to experience, this results in 86,354 remaining pairs. For Conscientiousness, this results in 87,514 remaining pairs; 105,033 pairs remaining for Extraversion; 93,607 for Agreeableness; and 98,823 pairs for Neuroticism. The varying numbers between the facets are a result of missing data for the personality scores. All sample sizes for the different facets are within at least two standard deviations from the mean sample size, making them comparable. Each pair of connected users on Facebook was assigned to one of three categories. For *same low* pairs, both users are low scorers; for *mixed* pairs, one user is a low scorer, while the other is a high scorer; for *same high* pairs, both users are high scorers (Table 1).

Table 2. Definition of low and high scorers for each personality facet

facet	low scorer	high scorer
Openness to Experience	not open	open
Conscientiousness	unconscientious	conscientious
Extraversion	introverted	extraverted
Agreeableness	disagreeable	agreeable
Neuroticism	emotionally stable	neurotic

We first analyzed the average distance to friends of Facebook users, depending on their own personality tercile score, but independently of the personality of their friends. This is to uncover any inherent tendencies of people of certain personality dispositions. Same pairs were then compared to mixed pairs, with the expectation of finding a significant difference in mean geographic distance, based on Hypotheses 1-5. Welch's t-test was used for all analyses. This is an alternative to the Student t-test, which is robust against unequal sample sizes and unequal variances.[32]

In light of the numerous tests conducted, we adopt a Bonferroni-corrected alpha-level of .0025, by dividing the usual alpha-level of .05 by the number of tests (20) carried out. The False Discovery Rate (FDR) was also checked with the Benjamini-Hochberg procedure.

Results

Using Welch's t-test, we found significant differences between low scorers and high scorers in terms of distance to friends for all facets (Table 3). Open people had friends living further away on average, compared to not-open individuals. Conscientious people tended to live further apart from their friends compared to unconscientious Facebook users. Agreeable users were also more likely to live further away from their friends compared to disagreeable users. Extraverted users, on the other hand, tended to live closer to their friends, compared to introverted users. Similarly, neurotic users had friends closer to them, on average, as compared to emotionally stable Facebook users.

Table 3. Mean distance (in km) between brokers and their friends for each facet, separated by high and low scorers

personality		**distance**		**Welch's t-test**		
facet	score	mean	SD	W	df	p-value
Open	high	451	925	641	586,022	<.0001
	low	394	857			
Consc	high	468	929	1664.31	592,275	<.0001
	low	376	848			
Extra	high	409	878	374.63	539,191	<.0001
	low	452	916			
Agree	high	432	901	35.27	647,325	<.0001
	low	419	891			
Neuro	high	412	882	97.34	670,783	<.0001
	low	433	899			

Openness to experience

For the Openness to experience facet (Hypothesis 1), the difference in geographic distance between the different types of connected pairs was significant, t(2,10424.72)=11.257, p<.0001 (Table 4). Pairs of same low users for Openness to experience (M=417.1km, SD=875.84) were significantly closer to each other in terms of geographic distance than users in mixed pairs (M=483.72km, SD=957.96). We found that the geographic distance was significantly higher for pairs of same high scoring pairs (M=501.79km, SD=991.61), compared to same low scoring pairs, but not compared to mixed pairs (p=.245).

This might be explained by the fact that not-open people do not tend to move around, and stay in close proximity to similar others, while open people tend to travel further away, and therefore are more spread out.

Conscientiousness

For Conscientiousness (Hypothesis 2), there was a significant effect of personality on distance among connected pairs, t(2,8593.677)=41.388, p<.0001 (Table 4). Pairs of same low

scoring users for Conscientiousness (M=398.69km, SD=923.74) and mixed pairs (M=438.88km, SD=910.57) did not differ significantly in terms of geographic distance (p=.038) when adopting a Bonferroni-corrected alpha-level of .0025. However, with a FDR-corrected p-value of .048 using the Benjamini-Hochberg procedure, this difference is statistically significant. Same high scoring pairs (M=549.61km, SD=1005.21) were significantly further apart than mixed pairs. Same low scoring and high scoring pairs also differed significantly, see Table 4.

Extraversion

For the Extraversion facet (Hypothesis 3), geographic distance differed significantly depending on the different pairs, t (2,11503.83) = 24.851, p<.0001 (Table 4). Same low scoring pairs on the Extraversion facet lived the furthest apart (M=545.03km, SD=998.05), compared to mixed pairs (M=464.79km, SD=926.9). Same high scoring pairs lived closest together (M=420.0km, SD=897.0), compared to mixed pairs, and same low scoring pairs. This provides support for Hypothesis 3: extraverts like to physically meet up with others to maintain their relationship, which might explain their closer proximity.

Agreeableness and Neuroticism

We found no support for Hypothesis 4 in regards to user pairs' geographic distance. There was no significant difference in geographic distance for pairs of connected users (t(2,11227.37)=0.731,p=.481) for the Agreeableness facet. We found no support for the Neuroticism facet (Hypothesis 5) either, t(2,11138.53)=2.652,p=.071).

Table 4. Welch's t-test results for openness to experience, conscientiousness, and extraversion.

Pair type	N	W	p-value
Openness to Experience			
Same low / mixed	7802.85	14.335	<.0001
Same high / same low	8490.417	21.201	<.0001
Conscientiousness			
Same high / mixed	17300.441	58.086	<.0001
Same high / same low	5723.205	27.38	<.0001
Extraversion			
Same low / mixed	7574.535	21.028	<.0001
Same high / mixed	19173.706	11.826	.001
Same high / same low	7947.726	49.127	<.0001

Discussion

The aim of this paper was to study the interaction of geographic distance and personality on the connection between Facebook users.

In accordance with our hypotheses, we found that geographic distance between a pair of users differed depending on their personality. Notably, we find that people who scored high on the Openness to experience, Conscientiousness, and Agreeableness were geographically further apart from their friends compared to people who scored low on these same facets. On the other hand, high scorers on the Extraversion and Neuroticism facet were found to be geographically closer to their friends, compared to low scorers on those facets.

To further investigate this effect, we compared connected pairs of users, based on their personality composition (same or mixed). We confirmed that conscientious and open pairs of users were indeed further apart than their low-scoring counterparts, as well as pairs with mixed personality scores on these facets.

It is important to note that most of these relationships have relatively small effect sizes. However, the relationship between personality and geographic distance, as well as the small effect sizes, are expected and consistent with previous work.[10,28,33]In addition, we have used conservative methods to control for family-wise error by applying Bonferroni corrections, and have checked False Discovery Rates (FDR) with the Benjamini-Hochberg procedure. A replication using a different sample would clarify whether pairs scoring low on Conscientiousness are indeed closer geographically than mixed pairs. Social media research often benefit from large samples that do not rely on large effect sizes to be replicable.[34]

Further work can focus on the reasons behind these effects. The static nature of the current dataset is restrictive. A more dynamic network approach, which follows people and their connections as they move, would be able to give further insight. Personality has been found to influence how we keep in touch with others, which shows the importance of such an approach.[35]

The current study does not take into account the goal behind user connections, which might range from genuine personal connections to opportunistic networking. Future research could seek to identify the impact of user motivation when maintaining a Facebook contact, on the interaction between personality homophily and geographic distance. However, it remains unclear the role quality, rather than quantity, of exchange plays in this case.

The quality of online relationships should be considered as well. Offline-first relationships are qualitatively the best, followed by mixed-mode relationships, and then online-only ones.[20] Most Facebook friendships fall in the offline category, with people having met on average 96% of their Facebook friends in person before connecting online.[36] Just like offline networks, social media connections vary greatly, with 21% of Facebook friends considered close friends. With an average of 400 friends, this equates to 80 close connections, which is the size of

a person's active network.[37] This is also in line with research finding that online and offline networks mirror each other in terms of the size of the different network layers.[38] Finally, with the exponential increase in the use of social media over the last decade, it is valuable to consider online networks as a worthy subject of study in their own right. Web-based studies are mostly as reliable and diverse as more traditional, offline studies, and greatly benefit from large sample sizes which can help to increase statistical power.[39]

Conclusion

Pairs of extraverted friends tended to live closer together than pairs of introverted or mixed friends. On the other hand, pairs of open friends tended to live further apart compared to not-open pairs. Conscientious users were also on average further apart from each other, compared to unconscientious ones. However, the snapshot nature of the data does not allow us to make any inferences on the reasons behind these effects. Future work will need to address how such relationships evolve with distance in a longitudinal context, which also takes into account other variables, such as frequency and quality of contact, which have been found to be essential for the maintenance of social ties. Personality homophily and how it relates to network structure remains an understudied phenomenon in the current literature. To the best of our knowledge, this study is one of the first to link personality homophily and geographic distance together.

Acknowledgments

Author Disclosure Statement

No competing financial interests exist.

References

1. Kossinets G, Watts DJ. Empirical analysis of an evolving social network. Science (New York, NY). 2006; 311:88–90.
2. Kossinets G, Watts DJ. Origins of Homophily in an Evolving Social Network. American Journal of Sociology. 2009; 115:405–50.
3. McPherson M, Smith-Lovin L, Cook JM. Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology. 2001; 27(1):415–44.
4. Sherman RA, Nave CS, Funder DC. Properties of persons and situations related to overall and distinctive personality-behavior congruence. Journal of Research in Personality. 2012; 46(1):87–101.
5. Goldberg LR. An alternative "description of personality ": The big-five factor structure. Journal of Personality and Social Psychology. 1990; 59(6):1216–29.
6. McCrae RR, Costa PT. Validation of the five-factor model of personality across instruments and observers. Journal of personality and social psychology. 1987; 52(1):81–90.
7. Gil de Zuniga H, Diehl T, Huber B, Liu J. Personality Traits and Social Media Use in 20 Countries: How Personality Relates to Frequency of Social Media Use, Social Media News Use, and Social Media Use for Social Interaction. Cyberpsychology, behavior and social networking. 2017; 20(9):540–52.
8. Selfhout M, Burk W, Branje S, Denissen J, van Aken M, Meeus W. Emerging Late Adolescent Friendship Networks and Big Five Personality Traits: A Social Network Approach. Journal of Personality. 2010; 78(2):509–38.
9. Balmaceda JM, Schiaffino S, Godoy D. How do personality traits affect communication

among users in online social networks? Online Information Review. 2013;38:136–53.

10. Noë N, Whitaker RM, Allen SM. Personality Homophily and the Local Network Characteristics of Facebook. In 2016 IEEE/ACM International Conference on the Advances in Social Network Analysis and Mining (ASONAM) San Francisco, CA, August 18-21: IEEE, pp. 386-393.
11. Selden M, Goodie AS. Review of the effects of Five Factor Model personality traits on network structures and perceptions of structure. Social Networks. 2017
12. Kalish Y. Bridging in social networks: Who are the people in structural holes and why are they there? Asian Journal of Social Psychology. 2008; 11:53–66.
13. Kalish Y, Robins G. Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure. Social Networks. 2006; 28(1):56–84.
14. Scellato S, Mascolo C. Distance matters: Geo-social metrics for online social networks. Proceedings of the 3rd conference on Online social networks. 2010;1–8.
15. Liben-Nowell D, Novak J, Kumar R, Raghavan P, Tomkins A. Geographic routing in social networks. Proceedings of the National Academy of Sciences. 2005; 102(33):11623–8.
16. Lambiotte R, Blondel VD, de Kerchove C, Huens E, Prieur C, Smoreda Z, et al. Geographical dispersal of mobile communication networks. Physica A: Statistical Mechanics and its Applications. 2008;387(21):5317–25.
17. Backstrom L, Sun E, Marlow C. Find Me If You Can: Improving Geographical Prediction with Social and Spatial Proximity. Proceedings of the 19th International Conference on World Wide Web. 2010; 61–70.

18. Barthélemy M. Spatial networks. Physics Reports. 2011; 499(1–3):1–101.

19. Deville P, Song C, Eagle N, Blondel VD, Barabási A-L, Wang D. Scaling identity connects human mobility and social interactions. Proceedings of the National Academy of Sciences. 2016; 113(26):7047–52.

20. Antheunis ML, Valkenburg PM, Peter J. The quality of online , offline , and mixed - mode friendships among users of a social networking site. Cyberpsychology: Journal of Psychosocial Research on Cyberspace. 2012; 6(3):1–13.

21. Graham LT, Gosling SD. Can the Ambiance of a Place be Determined by the User Profiles of the People Who Visit It? Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011; 145–52.

22. Jokela M, Bleidorn W, Lamb ME, Gosling SD, Rentfrow PJ. Geographically varying associations between personality and life satisfaction in the London metropolitan area. Proceedings of the National Academy of Sciences. 2015; 1–6.

23. Rentfrow PJ, Jokela M, Lamb ME. Regional Personality Differences in Great Britain. Plos One. 2015;10(3):e0122245.

24. Chorley MJ, Colombo GB, Allen SM, Whitaker RM. (2013) Visiting patterns and personality of foursquare users. In 2013 Third International Conference on Cloud and Green Computing (CGC). Karlsruhe, Germany, September 30 - October 2: IEEE, pp. 271-276.

25. Chorley MJ, Whitaker RM, Allen SM. Personality and location-based social networks. Computers in Human Behavior. 2015; 46: 45-56.

26. Noë N, Whitaker RM, Chorley, MJ, Pollet TV. Birds of a feather locate together? Foursquare checkins and personality homophily. Computers in Human Behavior 2016; 58:

343-353.

27. Goldberg LR. An alternative “description of personality”: the big-five factor structure. Journal of personality and social psychology. 1990; 59(6):1216–29.

28. Ross C, Orr ES, Sisic M, Arseneault JM, Simmering MG, Orr RR. Personality and motivations associated with Facebook use. Computers in Human Behavior. 2009; 25(2):578–86.

29. Hughes DJ, Rowe M, Batey M, Lee A. A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. Computers in Human Behavior. 2012; 28(2):561–9.

30. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. American Psychologist. 2015; 70(6):543.

31. Amichai-Hamburger Y, Vinitzky G. Social network use and personality. Computers in Human Behavior. 2010; 26(6):1289–95.

32. Ruxton GD. The unequal variance t-test is an underused alternative to Student’s t-test and the Mann-Whitney U test. Behavioral Ecology. 2006;17(4):688–90.

33. Butcher JN, Graham JR, Ben-Porath YS. Methodological Problems and Issues in MMPI, MMPI-2, and MMPI-A Research. Psychological Assessment. 1995;7(3):320–9.

34. Matz SC, Gladstone JJ, Stillwell D. In a World of Big Data, Small Effects Can Still Matter. Psychological Science. 2017

35. Huang HC, Cheng TCE, Huang WF, Teng CI. Who are likely to build strong online social networks? The perspectives of relational cohesion theory and personality theory. Computers in Human Behavior. 2018; 82:111–23.

36. Manago AM, Taylor T, Greenfield PM. Me and my 400 friends: The anatomy of college students' facebook networks, their communication patterns, and well-being. Developmental Psychology. 2012; 48(2):369–80.

37. Hill RA, Dunbar RIM. Social network size in humans. Human Nature. 2003; 14(1):53–72.

38. Dunbar RIM, Arnaboldi V, Conti M, Passarella A. The structure of online social networks mirrors those in the offline world. Social Networks. 2015; 43:39–47.

39. Gosling SD, Vazire S, Srivastava S, John OP. Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. American Psychologist. 2004; 59(2):93–104.